JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 76, 222–229 (1980)

Generalization of a Theorem by Hardy, Littlewood, and Pólya

ERNST RUCH

Institut für Quantenchemie der Freien Universität Berlin

RUDOLF SCHRANNER

Max-Planck-Institut für Biophysikalische Chemie Göttingen

AND

THOMAS H. SELIGMAN

Instituto de Fisica, Universidad Nacional Autonoma de Mexico

Submitted by G.-C. Rota

A theorem by Hardy, Littlewood, and Pólya [1], in the present paper referred to as HLP-theorem, applies to the set of nonnegative vectors with a given sum of the coefficients, i.e., to sets $\bar{\mathbb{R}}^n = \{p \mid p \in \mathbb{R}^n, \sum_1^n p_i = \text{const.}, p_i \geqslant 0\}$ and states the equivalence of a system of inequalities with the action of bistochastic matrices

$$\forall r \leqslant n: \quad \sum_{i=1}^{r} p'_{\pi'(i)} \leqslant \sum_{i=1}^{r} p_{\pi(i)} \quad \Leftrightarrow \quad \exists B \in \mathbb{B}: \quad p' = Bp. \tag{1}$$

$\mathbb{B}$ denotes the semigroup of bistrochastic $n \times n$ matrices and $\pi(i)$, $\pi'(i)$ are two index permutations such that $p_{\pi(1)} \geqslant p_{\pi(2)} \geqslant \cdots \geqslant p_{\pi(n)}$ and $p'_{\pi'(1)} \geqslant p'_{\pi'(2)} \geqslant \cdots \geqslant p'_{\pi'(n)}$.

The inequalities on the left side of (1) represent an order relation between equivalence classes in the vector set $\bar{\mathbb{R}}^n$, equivalence being defined between vectors with permuted coefficients. The set of these classes according to (1) is partially ordered and was proved to be a lattice [2]. The lattice structure applies also to Young diagrams and, correspondingly generalized, to nonnegative functions of various types. We found this structure of relevance in many seemingly unrelated contexts such as representation theory of the symmetric group [2, 3], graph theory [4], stereochemistry [2], information theory [5] and statistical mechanics [3, 7]. Applications related to the last two disciplines of physics, in particular the formulation of irreversibility in nonisolated systems

0022-247X/80/070222-08$02.00/0

and the definition of appropriate information theoretical concepts led to an order relation of a similar type, which applies to classes in the set of pairs of non-negative vectors or functions. The valuable context with bistochastic matrices according to (1), by analogy suggests a corresponding theorem referring to stochastic matrices which appears to be the natural generalization of the HLP-theorem. We proved the generalized theorem in "a physicist's way" using Dirac δ-functions in Ref. [8]. The aim of the present paper is to present the proof in precise mathematical terms, without our original restrictive assumptions.

One takes advantage from using concepts, which appeared to be valuable for the mentioned applications. All vectors, which pairwise satisfy the left side of (1) in the particular form of equalities were called "mixing equivalent" and the property common to mixing equivalent vectors their "mixing character." $m[p]$ was used to denote the mixing character of a vector p and the relation $m[p'] \succ m[p]$ may be read as "p' is more mixed than p" or "the mixing character of p' is larger than the mixing character of p." Using an equivalent form of the inequalities in (1) which is easily confirmed (comp. [6, 8]), we present the "more mixed relation" as follows:

DEFINITION 1.

$$m[p'] \succ m[p] \quad \Leftrightarrow \quad \forall\, l \in \mathbb{R}^+ : \sum_{i=1}^{n} | p'_i - l | \leqslant \sum_{i=1}^{n} | p_i - l |.$$

Now, the HLP-theorem takes the following concise form.

THEOREM 1 (HLP-theorem).

$$m[p'] \succ m[p] \quad \Leftrightarrow \quad \exists\, B \in \mathbb{B} : p' = Bp.$$

The parameter l in Definition 1 may be replaced by l/n with $1/n$ denoting the components of the uniform vector in $\bar{\mathbb{R}}^n$. Increasing mixing character means therefore increasing closeness to the uniform vector or, as pointed out in [6, 8], decreasing mixing distance from maximal mixing. In consequence of and in agreement with physical phenomena and their statistical interpretation, the concept of mixing distance should also be applicable to an arbitrary pair of vectors in $\bar{\mathbb{R}}^n$. A correspondingly generalized order relation follows from physically stringent arguments [3] and reads with $d[p/q]$ denoting "the mixing distance of p from q" as follows:

DEFINITION 2.

$$d[p'/q'] \prec d[p/q] \quad \Leftrightarrow \quad \forall\, l \in \mathbb{R}^+ : \sum_{i=1}^{n} | p'_i - l q'_i | \leqslant \sum_{i=1}^{n} | p_i - l q_i |.$$

Now, recalling that bistrochastic matrices are particular stochastic matrices that leave the uniform vector invariant we present the following Theorem 2 connecting the mixing distance order relation with the action of stochastic matrices as an obvious conjecture:

THEOREM 2 (generalized HLP-theorem).

$$d[p'/q'] \prec d[p/q] \quad \Leftrightarrow \quad \exists\, S \in \mathbb{S}: \begin{cases} p' = Sp \\ q' = Sq \end{cases}$$

with $\mathbb{S}$ *denoting the set of stochastic* $n \times n$ *matrices.*

This theorem, though concerned with discrete mathematics, shall be proved for the more general case of functions in $L^1([0, 1])$ using the means of functional analysis.

HLP-THEOREM FOR L^1-FUNCTIONS ACCORDING TO RYFF [9]

We need a summary of facts on the mixing character of functions, in order to present Ryff's result in terms which we shall use subsequently. Let us call a positive function $p \in L^1([0, 1])$ with $\int_0^1 p(t)\, dt = 1$ a (probability–)density. (Note that functions $p \in L^1$ differing on subspaces of measure zero in $[0, 1]$ are taken to be equivalent.) The integral $P(t) = \int_0^t p(\tau)\, d\tau$ sometimes called (probability-)distribution function, is a continuous and monotonous function on $[0, 1]$. Associated with any given density $p \in L^1([0, 1])$ there is a density p^* which is monotonously decreasing and satisfies the relation $p(t) = p^*(\varphi(t))$, where φ is a measure preserving transformation on the interval $[0, 1]$ (comp. [9]). p^* was called "decreasing arrangement of p^1" by Ryff; we prefer the notation "diagram density" emphasizing the analogy with the arrangement of rows in a Young diagram. The mixing character of a density p permits a variety of equivalent definitions, which are straightforward translations from corresponding versions in the case of positive vectors and Young diagrams (comp. [3, 6, 8]). We shall use the equivalence of the following statements.

(a) $m[p] \succ m[p']$,

(b) $\int_0^1 | p(t) - l \,|\, dt \leqslant \int_0^1 | p'(t) - l \,|\, dt \quad \forall\, l \in \mathbb{R}^+$,

(c) $\int_0^1 (p(t) - l)^+\, dt \leqslant \int_0^1 (p'(t) - l)^+\, dt \quad \forall\, l \in \mathbb{R}^+$ with $x^+ = \max\{x, 0\}$,

(d) $P^*(t) \leqslant P'^*(t)$, where $P^*(t) = \int_0^t p^*(\tau)\, d\tau$.

$P^*(t)$ corresponds to the partial sums $\sum_1^r p_{\pi(i)}$ in (1), which were used in Ref. [2] for proving that the order relation (Definition 1) defines a lattice. The same arguments apply to $L^1([0, 1])$ densities. Therefore the classes of mixing equivalent

$L^1([0, 1])$ densities establish a lattice with union and intersection defined as follows:

$$\sup(m[p], m[q]) = m[s], \qquad \text{where} \quad s(t) = \frac{d}{dt}\max(P^*(t), Q^*(t)),$$

$$\inf(m[p], m[q]) = m[i], \qquad \text{where} \quad i(t) = \frac{d}{dt}\min(P^*(t), Q^*(t)).$$

The functions s and i represent proper probability densities of $L^1([0, 1])$ since the derivatives in the above expressions may be ill defined at most on a subset of measure zero. Note that i is a diagram density whereas s, if constructed as given above, is not.

A linear operator B on $L^1([0, 1])$, according to Rota's definition [10], is called bistochastic if it has the following properties:

(α) B is positive,
(β) $\int_0^1 (Bf)(t)\, dt = \int_0^1 f(t)\, dt \;\forall f > 0$,
(γ) $B1 = 1$.

With p, p' denoting densities and $\mathbb{B}$ denoting the set of bistochastic operators in $L^1([0, 1])$ instead of normalized vectors and $n \times n$ matrices, respectively, in $\overline{\mathbb{R}}^n$, Theorem 1 as given above represents Ryff's result in terms of the more mixed relation.

The Generalized HLP-Theorem

The following discussion includes the discrete case, since vectors in $\overline{\mathbb{R}}^n$ may be considered as step densities of $L^1([0, 1])$.

The mixing distance order relation according to Definition 2, extended to $L^1([0, 1])$ densities allows a variety of equivalent definitions which are the respective analogon of these for the mixing character and are proved in Refs. [5, 8]. Here, we need the equivalence of the following statements.

(a) $d[p'/q'] \prec d[p/q]$,
(b) $\forall l \in \mathbb{R}^+$: $\int_0^1 | p'(t) - lq'(t)|\, dt \leqslant \int_0^1 | p(t) - lq(t)|\, dt$,
(c) $\forall l \in \mathbb{R}^+$: $\int_0^1 (p'(t) - lq'(t))^+\, dt \leqslant \int_0^1 (p(t) - lq(t))^+\, dt$.

The concept of mixing distance is associated with the formation of classes of mixing distance equivalent pairs defined by

$$(p, q) \sim (p', q') \quad \Leftrightarrow \quad d[p/q] = d[p'/q'].$$

Substituting the parameter $1/l$ for l in (b) or (c) and multiplying subsequently both sides of the inequalities with l leads to the following lemma.

LEMMA.

$$d[p/q] \prec d[p'/q'] \quad \Leftrightarrow \quad d[q/p] \prec d[q'/p'].$$

It follows that $d[p/q]$ and $d[q/p]$ are either identical or incomparable.

The characterization of the mixing distance order relation in terms of operations is based on the following definition of stochastic operators $L^1([0, 1])$.

A linear operator S is called stochastic operator in $L^1([0, 1])$ if

(α) S is positive,

(β) $\int_0^1 (Sf)(t)\,dt = \int_0^1 f(t)\,dt \quad \forall f \in L^1([0, 1])$.

With $\mathbb{S}$ denoting the set of such operators and p, q, p', q' representing densities, the generalized HLP-theorem for $L^1([0, 1])$ densities is formally identical with Theorem 2, which we prove now with reference to said densities.

The proof of

$$d[p'/q'] \prec d[p/q] \quad \Leftarrow \quad \exists\, S \in \mathbb{S}: \begin{cases} p' = Sp \\ q' = Sq \end{cases}$$

is straightforward since the definition of the mixing distance relation given by (b) allows applying the triangle relation for absolute values.

$$\int_0^1 |p'(t) - lq'(t)|\,dt = \int_0^1 |(Sp)(t) - l(Sq)(t)|\,dt \leqslant \int_0^1 (S\,|\,p - lq\,|)(t)\,dt$$
$$= \int_0^1 |\,p(t) - lq(t)|\,dt.$$

The proof of

$$d[p'/q'] \prec d[p/q] \quad \Rightarrow \quad \exists\, S \in \mathbb{S}: \begin{cases} p' = Sp \\ q' = Sq \end{cases}$$

shall be given in three steps, (a) by assuming that q and q' are strictly positive and (b) and (c) by discussing the problem which arises from zeros of the densities q and q'.

(a) We assume that q is strictly positive and note that the distribution function $Q(t) = \int_0^t q(\tau)\,d\tau$ in consequence is strictly increasing and therefore invertible. Thus, $Q^{-1}(t) = \sigma$ is a transformation of variables in the interval $[0, 1]$ and allows defining a relative density $\pi_{p/q}(t)$ in $L^1([0, 1])$ as follows:

$$\pi_{p/q}(t) = \frac{p(Q^{-1}(t))}{q(Q^{-1}(t))}\,.$$

Now, the equation

$$\int d\sigma\,|\,p(\sigma) - lq(\sigma)| = \int dt \left| \frac{p(Q^{-1}(t))}{q(Q^{-1}(t))} - l \right|$$

entails the equivalence (2).

$$d[p'/q'] \prec d[p/q] \Leftrightarrow m[\pi_{p'/q'}] \succ m[\pi_{p/q}] \qquad \text{with} \quad q > 0, \quad q' > 0. \qquad (2)$$

Since any density of $L^1([0, 1])$ can be written as a relative density $\pi_{p/q}$ with a fixed $q > 0$ there are linear and bijective mappings from $L^1([0, 1])$ onto $L^1([0, 1])$ representable by stochastic operators T_q, $T_{q'}$ and T_q^{-1}, $T_{q'}^{-1}$ such that

$$T_q p = \pi_{p/q}, \qquad T_{q'} p' = \pi_{p'/q'},$$
$$T_q^{-1}\pi_{p/q} = p, \qquad T_{q'}^{-1}\pi_{p'/q'} = p'.$$

Because of the HLP-theorem for $L^1([0, 1])$ densities the right side of (2) can be written

$$\pi_{p'/q'} = B\pi_{p/q}, \qquad \text{where } B \text{ is a bistochastic operator.}$$

Noting that $T_{q'}^{-1}BT_q$ is stochastic, we obtain

$$p' = T_{q'}^{-1}BT_q p,$$
$$q' = T_{q'}^{-1}1 = T_{q'}^{-1}B1 = T_{q'}^{-1}BT_q q.$$

Therefore the relation

$$d[p'/q'] \prec d[p/q] \Rightarrow \begin{cases} p' = Sp \\ q' = Sq \end{cases} \qquad \text{with} \quad S \in \mathbb{S}$$

is proved assuming strictly positive reference densities q and q'.

(b) We take a step towards the general situation in reducing the condition of strictly positive reference densities q and q' to the weaker assumption $q(t) = 0 \Rightarrow p(t) = 0$ and $q'(t) = 0 \Rightarrow p'(t) = 0$. In this case, it is correct to agree upon fixing $q(t) = 0 \Rightarrow \pi_{p/q}(Q(t)) = 0$ and $q'(t) = 0 \Rightarrow \pi_{p'/q'}(Q'(t)) = 0$. Thus, both the expressions $\pi_{p/q}$ and $\pi_{p'/q'}$ become well defined and normalized L^1 densities on $[0, 1]$, and the result of the proof in (a) follows also under the weaker presuppositions of part (b).

(c) We handle the general case discussing the l-dependence of the integrals $J_l = \int_0^1 (p(t) - lq(t))^+ \, dt$ by means of a new parameter s, that is related to l as follows:

$$s = \frac{\alpha l + \beta}{\beta l + \alpha}, \qquad l = \frac{\alpha s - \beta}{\alpha - \beta s} \qquad \text{with} \quad 0 < \beta < \alpha < \alpha + \beta = 1.$$

Thereby the nonnegative range $0 \leqslant l$ is bijectively mapped onto $\beta/\alpha \leqslant s \leqslant \alpha/\beta$ and exclusively negative values of l correspond to $s < \beta/\alpha$ and $\alpha/\beta < s$. A simple calculation shows, that for each of the ranges $0 \leqslant s < \beta/\alpha$ and $\alpha/\beta < s$ the following expression is, respectively, definite.

$$(\alpha p(t) + \beta q(t)) - s(\beta p(t) + \alpha q(t))$$
$$= \frac{\alpha^2 - \beta^2}{\beta l + \alpha}(p(t) - lq(t)) \begin{cases} \geqslant 0 & \text{for } 0 \leqslant s < \dfrac{\beta}{\alpha}, \\ \leqslant 0 & \text{for } \dfrac{\alpha}{\beta} < s. \end{cases}$$

Therefore we have for any p and q:

$$\forall s \in \left\{s \,\middle|\, 0 \leqslant s < \frac{\beta}{\alpha}\right\}: \int_0^1 [(\alpha p(t) + \beta q(t)) - s(\beta p(t) + \alpha q(t))]^+ \, dt = 1 - s,$$
$$\forall s \in \left\{s \,\middle|\, \frac{\alpha}{\beta} < s\right\}: \int_0^1 [(\alpha p(t) + \beta q(t)) - s(\beta p(t) + \alpha q(t))]^+ \, dt = 0.$$

From this, one concludes the following equivalence

$$\forall s \in \mathbb{R}^+: \int_0^1 [(\alpha p'(t) + \beta q'(t)) - s(\beta p'(t) + \alpha q'(t))]^+ \, dt$$
$$\leqslant \int_0^1 [(\alpha p(t) + \beta q(t) - s(\beta p(t) + \alpha q(t))]^+ \, dt$$
$$\Leftrightarrow \forall l \in \mathbb{R}^+: \int_0^1 (p'(t) - lq'(t))^+ \, dt \leqslant \int_0^1 (p(t) - lq(t))^+ \, dt,$$

or, in concise form,

$$d\left[\frac{\alpha p' + \beta q'}{\beta p' + \alpha q'}\right] \prec d\left[\frac{\alpha p + \beta q}{\beta p + \alpha q}\right] \quad \Leftrightarrow \quad d[p'/q'] \prec d[p/q].$$

Our Lemma indicates that this equivalence remains true if densities and reference densities are interchanged; in other words, our result is valid for $\alpha \lessgtr \beta$, $\alpha + \beta = 1$. For all $\alpha \neq 0$, $\alpha + \beta = 1$ on the other hand, we have the following conclusions:

$$\beta p(t) + \alpha q(t) = 0 \quad \Rightarrow \quad \alpha p(t) + \beta q(t) = 0,$$
$$\beta p'(t) + \alpha q'(t) = 0 \quad \Rightarrow \quad \alpha p'(t) + \beta q'(t) = 0,$$

where p, q, p', q' are arbitrary densities without restricting conditions. Thus, following the reasoning of the proof, part (b), we conclude:

$$d[p'/q'] \prec d[p/q] \Rightarrow \exists\, S \in \mathbb{S}: \begin{cases} \alpha p' + \beta q' = S(\alpha p + \beta q) \\ \beta p' + \alpha q' = S(\beta p + \alpha q). \end{cases}$$

With $\alpha \neq \beta$, it follows that

$$p' = Sp,$$
$$q' = Sq.$$

This result completes the proof of the conjectured HLP-theorem for $L^1([0, 1])$ densities and in particular for vectors in $\overline{\mathbb{R}}^n$.

REFERENCES

1. G. H. HARDY, J. E. LITTLEWOOD, AND G. PÓLYA, "Inequalities," 2nd Ed. Cambridge Univ. Press, London/New York, 1973.
2. E. RUCH AND A. SCHÖNHOFER, *Theor. Chim. Acta* **19** (1970), 225.
3. E. RUCH, *Theor. Chim. Acta* **38** (1975), 167.
4. E. RUCH AND I. GUTMAN, to be published.
5. E. RUCH AND B. LESCHE, *J. Chem. Phys.* **69** (1978), 393.
6. E. RUCH AND A. MEAD, *Theor. Chim. Acta* **41** (1976), 95.
7. A. MEAD, *J. Chem. Phys.* **66** (1977), 459.
8. E. RUCH, R. SCHRANNER AND T. H. SELIGMAN, *J. Chem. Phys.* **69** (1978), 386.
9. J. V. RYFF, *Trans. Amer. Math. Soc.* **117** (1965), 92; *Proc. Amer. Math. Soc.* **18** (1967), 1026.
10. G.-C. ROTA, *Bull. Amer. Soc.* **68** (1962), 95.